June 9, 2000

Donald Barnes, Ph.D.
Science Advisory Board (1400A)
USEPA
Washington, DC 20460

Dear Dr. Barnes:

Attached are comments from the American Crop Protection Association on the SAB Executive Committee Review Draft Report "Report Of The Joint Subcommittee On data From Testing Of Human Subjects". Also included with this communication are two attachments, expert papers that ACPA had provided to the Joint SAB/SAP Subcommittee on January 14, 2000, "The Substantial Power of Human Study Data to Contribute to the Characterization of NOELs for Cholinesterase Inhibition in Humans: A Statistical Analysis of Recent Studies" and "An Overview of the Biological Significance of variations in Blood Cholinesterase Measurements." All of this information is being provided to you as a Word 97 documents.

Please furnish copies of all these documents to the members of the SAB Executive Committee for their review prior to the teleconference meeting that is to be held on June 16, 2000. A hard copy is also being sent for your files by FAX.

Thank you for your efforts.

Sincerely yours,

Angelina J. Duggan, Ph.D.
ACPA Director of Science Policy

Cc: Larry Dorsey, EPA
Samuel Rondberg, EPA

American Crop Protection Association Comments
The Report Of The Joint SAB/SAP Subcommittee On Data From Testing Of Human Subjects

Executive Committee Review Draft Report
May 31, 2000

Submitted by Angelina J. Duggan, Ph.D.
June 9, 2000

ACPA is pleased that the Joint SAB/SAP Subcommittee has endorsed the importance of conducting human studies in providing valuable scientific information for biomonitoring and elucidating the absorption, excretion, pharmacokinetics and metabolism of pesticides in humans. We are also pleased that the Subcommittee has compared the benefits of pesticides to pharmaceuticals and recognizes the essential role of these products in food production and in public health.

However, ACPA does not agree with the Subcommittee's general conclusion that future human volunteer studies should not be conducted with the intent to determine no effect or lowest effects levels or to establish references doses. The Subcommittee lists a number of arguments to justify their position, in particular that the NOEL used may not be the most sensitive one, that acute toxicity testing does not reflect the potential to cause long-term effects, and that data from healthy adult volunteers do not reflect processes in infants and children. None of these arguments are unique to human studies. They apply equally or in some cases more strongly to animal studies, and, if used as a basis for a decision not to accept human data, would therefore invalidate the whole concept of risk assessment that the agency is currently using for pesticides. ACPA considers it unscientific and unethical not to consider human studies in the same way as the rest of the database. It would be entirely illogical not to use human data for a specific purpose, provided they are scientifically sound and have been obtained in compliance with national and international guidelines concerning the conduct of human studies, for example, the Helsinki Convention and the Common Rule. Moreover, we contend that it is ethically and morally wrong to ignore studies that have already been conducted if the data adhered to sound scientific and ethical standards and had been subjected to the oversight of an Institutional Review Board (IRB).

There cannot be an arbitrary line for the application of human pesticide data in determining the overall benefits and risks to human populations that could be potentially exposed to these products. Subjective constraints should not be placed on the value or use of existing or future human testing data. For risk assessment purposes, human volunteer testing information, gained under carefully controlled biomedical conditions and with informed consent by participants, enhances and complements the information already gained from the animal laboratory tests.

In fact overall, it appears that the Subcommittee's recommendations are generally in agreement with ACPA's position on the use and benefits of human testing data:

> "… in the case of pesticides, a broader population is potentially exposed and not monitored for health effects. This situation is a powerful arguilent for the conduct of controlled exposure studies to better understand the effects of low level exposures. Otherwise, the populace is our controlled exposure study."
>
> "Human volunteer studies could be appropriate when there are significant data gaps and such studies could provide a more accurate risk assessment."
>
> "Human volunteer studies could be appropriate for pesticides, which are not yet on the market, i.e., new products.".
>
> "… from a toxicological standpoint, it is inappropriate to consider oral dosing any differently from the other two possible routes of human exposure to pesticides, e.g., inhalation and dermal exposures."

ACPA believes that several of the issues raised by the Subcommittee are not relevant and confusing to the key issue that is, the appropriate use of human testing data to refine risk assessments. The primary purpose of determining acute human effects data has been to reduce the interspecies uncertainty factor inherent from extrapolating animal cholinesterase activity to humans. This has been especially important for risk assessments based on animal cholinesterase measurements. We advocate that human testing be conducted <u>only</u> in healthy and consenting adults. The intraspecies safety factors compensate for the obvious fact that this type of testing cannot and will not be conducted in children or other vulnerable populations. Thus it is an not valid to conclude that human studies conducted in adults are not appropriate for assessing risk in children. Moreover, chronic and developmental effects are adequately addressed by the animal toxicology studies, and, when warranted, EPA will use these endpoints in risk assessment rather than cholinesterase inhibition.

ACPA recognizes that the Subcommittee has provided some useful guidance to EPA on the process for conducting human volunteers studies. This has included recognizing the critical role of the IRB in providing technical and scientific oversight <u>before</u> such evaluations are initiated and recommending that EPA undertake a more active role in the oversight planning process of these studies. The SAB Executive Committee should be aware that the contract laboratories, both domestic and abroad that have already conducted human volunteer studies (either those studies pending review or those that have been recently been reviewed by EPA) have complied with the scientific and ethical IRB review process that includes informed consent of the human volunteers.

ACPA supports the Subcommittee recommendation that the Agency organize a workshop to provide a thorough scientific review of the statistical power issue for human study design in order to provide guidance for future studies. We also recommend that the statistical power workshop cover both animal testing as well human testing since EPA currently establishes no effect and low effect levels on the basis of small numbers of test animals. The SAB Executive Committee should also be aware that the contract laboratories, both domestic and foreign, that have conducted human volunteer studies, either those studies pending review or those that have been recently been reviewed by EPA, have provided assessments of the statistical power of these studies. On January 14, 2000, ACPA submitted to the Subcommittee two expert papers, "The Substantial Power of Human Study Data to Contribute to the Characterization of NOELs for Cholinesterase Inhibition in Humans: A Statistical Analysis of Recent Studies" and "An Overview of the Biological Significance of variations in Blood Cholinesterase Measurements." Our findings concluded that the recent human studies (9 in all) have sufficient statistical power in excess of 90% to detect a 20% or higher cholinesterase inhibition. The WHO/FAO Joint Meeting on Pesticide Residues (JMPR) regards this as the threshold level for biologically significant cholinesterase inhibition. Since the Subcommittee has not referred to these ACPA papers in their draft report, we have included both of these papers with these comments. Additionally, we request that these papers be considered in the discussions of a future EPA workshop on statistical consideration.

An Overview of the Biological Significance of Variations in Blood Cholinesterase Measurements

William L. Chen, Ph.D.
Dow AgroSciences, LLC

Edward C. Gray
Jellinek, Schwartz & Connolly, Inc.

Martin F. Wilks, M.D., Ph.D.
Zeneca Agrochemicals

Abstract

For a number of organophosphate (OP) and carbamate insecticides, EPA uses no-observed-effect levels (NOELs) from studies of inhibition of blood cholinesterase (ChE) activity in deciding what exposure levels are allowable. While it is important to be able to detect large changes in blood ChE activity, small changes have no biological significance. We discuss the background blood ChE variability in unexposed humans. We also present the position of a number of authoritative individuals and organizations on biological significance of blood ChE inhibition. The position of the WHO/FAO Joint Meeting on Pesticide Residues (JMPR) is that statistically significant inhibition of 20% or more is regarded as biologically significant while inhibition of less than 20% must be reviewed on a case-by-case basis to determine whether it is of biological or significance. Thus, recent cholinesterase inhibition studies in humans are very useful even if they cannot detect very low percentages of inhibition, because they <u>can</u> detect biologically meaningful inhibition levels.

Introduction

The power of a study of a given size to detect a difference between dosed and placebo groups is a function of, among other things, the size of the difference in the response variable that one needs to detect for a particular event or effect. It is generally understood that large groups are needed in order to conclude reliably that a true small difference between tested and control groups was not missed in a test in which no effect was observed. However, it should be emphasized that at least with respect to blood ChE levels, small differences are neither biologically relevant nor relevant to human risk assessment. As stated in the accompanying paper by Robert L. Sielken and Larry Holden, it is critical to distinguish between the statistical and biological significance of observed events. Statistical and biological significance are not necessarily related, and it is not appropriate to use statistical significance to imply biological significance.

Variability of Human ChE Measurements and Their Implications for Risk Assessment

It is universally recognized that healthy individuals' blood ChE levels vary considerably over relatively short periods of time, and that these variations do not appear to cause adverse effects. Fluctuations of 13-25% for red blood cell acetylcholinesterase (RBC AChE) activity in an unexposed individual have been reported (Hayes, 1982). Due in part to this natural variability in RBC AChE activity, it is difficult to determine the degree of change that may be interpreted with confidence as inhibition as a result of pesticide exposure, rather than random variation (Hayes, 1982; Lotti, 1995). Most toxicologists do not treat small differences as biologically relevant.

The International Programme on Chemical Safety stated in its review of organophosphates (IPCS, 1986) that

> [I]t has been estimated that the coefficient of variation for AChE activity in samples from an individual is 8 - 11%, and that a decrease of 23% below pre-exposure level may, therefore, be considered significant. If the average of several pre-exposure values were available, then a decrease of 17% would be significant . . . Depressions of AChE or ChE in excess of 20 - 25% are considered diagnostic of exposure, but not, necessarily, of hazard. Depressions of 30 - 50% or more are considered indicators for removal of an exposed individual from further contact with the pesticides until levels return to normal.

The review of organophosphates published in Hayes and Laws' *Handbook of Pesticide Toxicology* (Gallo and Lawryk, 1991) considered that only changes of greater than 30% for plasma ChE or 20% for RBC ChE can be recognized with certainty as not due to normal variation. In Ballantyne and Marrs' authoritative book *Clinical and Experimental Toxicology of Organophosphates and Carbamates*, ChE depression of 20 - 25% is considered to reflect exposure, but not hazard (Lewinsohn, 1992).

Of particular importance are the views of the Joint Meeting on Pesticides Residues (JMPR), a FAO/WHO organization that recommends maximum residue levels (MRLs, the international pesticides tolerances) under the UN's Codex Alimentarius system. The JMPR's views on what levels of red blood cell ChE inhibition[1] are of biological (and thus regulatory) significance are as follows (FAO/WHO 1998):

> Regulatory agencies have traditionally used various thresholds, such as 10% inhibition, 20% inhibition, or any statistically significant inhibition, in defining biologically significant depression of enzyme activity. The Meeting considered that statistically significant inhibition by 20% or more represents a clear toxicological effect and any decision to dismiss such findings should be justified. The Meeting also agreed that statistically significant inhibition of less than 20% or statistically insignificant inhibition above 20% indicate that a more detailed analysis of the data should be undertaken. The toxicological significance of these findings should be determined on a case-by-case basis. Considerations affecting such determinations include *inter alia* the shape or slope of the dose-response curve, assay variability, and correlation with clinical signs.

Thus, there is no single measure by which to gauge the biological significance of ChE inhibition. However, there is general international consensus that a statistically significant decrease of 20% or greater is of regulatory significance, while decreases of less than 20% should be treated on a case-by-case basis. It follows that if human studies, like those recently conducted on OPs and carbamates, are sufficiently powerful to detect differences in the range of 15% or 20%, it would be incorrect to conclude that human studies always lack enough power for use in NOELS.

The accompanying paper by Sielken and Holden shows that the nine recently performed studies the authors analyzed are sufficiently powerful to detect such changes reliably. Accordingly, the Subcommittee cannot justifiably conclude that human studies are never sufficiently powerful to justify their use in establishing NOELs.

Other Factors That Increase Regulatory Confidence

The Subcommittee should keep in mind that the various pesticide studies in animals and humans can and should be viewed as complementary. It is not just the results of statistical power tests applied to individual data sets that tell us how confident we can be about the correctness of a conclusion drawn from a study. A series of consistent results within a study can lend confidence to the correctness of the overall conclusion, as can the consistent results of a set of various statistical analyses (e.g., analyses of individuals' variations from baseline over time and analyses of comparisons of group and control means at any given time). Different studies with consistent results likewise give regulators more confidence in the readings of each study individually. One presumes that this is a major reason why EPA, FDA, and other international regulatory agencies do not require animal studies to be larger than they are. Chronic and sub-chronic toxicity studies in dogs employ dose groups of no more than four animals per group. In fact, the number of animals used in these studies have been reviewed and are deemed appropriate by EPA, FDA, and other international regulatory authorities.

References

Gallo MA, Lawryk NJ (1991). Organic Phosphorus Pesticides. In: Hayes WJ, Laws ER (eds.), *Handbook of Pesticide Toxicology*, Vol. 2, pp. 917 - 1123, San Diego: Academic Press.

Hayes W (1982). *Pesticide studies in man.* Baltimore, MD: Williams & Wilkins.

FAO/WHO (1998), A Report of the FAO/WHO Joint Meeting of Pesticide Residues (JMPR), Section 2.14, Interpretation of Cholinesterase Inhibition, Rome 21-30 September 1998, pp 17-20.

IPCS International Programme on Chemical Safety (1986). *Organophosphorus Insecticides: A General Introduction.* Environmental Health Criteria 63, Geneva: World Health Organisation.

Lewinsohn HC (1992). General considerations on workplace safety for anticholinesterases. In: Ballantyne B, Marrs TC (eds.), *Clinical and Experimental Toxicology of Organophosphates and Carbamates*, pp. 399 - 413, Oxford: Butterworth Heinemann.

Lotti M (1995). Cholinesterase inhibition: Complexities in interpretation. *Clin. Chem*. 41: 1814-1818.

The Substantial Power of Human Study Data to Contribute to the Characterization of NOELs for Cholinesterase Inhibition in Humans: A Statistical Analysis of Recent Studies

Robert L. Sielken Jr., Ph.D., and Larry R. Holden, Ph.D.
JSC Sielken, 3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 409-846-5175; Fax: 409-846-2671; E-mail: SielkenINC@aol.com

January 14, 2000

Abstract

Data from placebo (control) subjects and baseline data from treatment subjects in nine recent studies that measured red blood cell (RBC) and plasma cholinesterase levels in humans were provided to JSC Sielken to facilitate the quantitative evaluation of the power of recent human studies to detect biologically significant levels of cholinesterase inhibition. The analyses in this report demonstrate that the study designs currently employed for human RBC and plasma ChE evaluation are all quite powerful for mean inhibition levels in the range of 15-25% below placebo levels. For example, the recent human studies evaluated herein have power usually substantially in excess of 90% to detect a 20% or higher level of cholinesterase inhibition. The statistical methodology used in the power calculations is described in the Appendix.

Introduction

The primary objective of the human studies of cholinesterase inhibition that we examined was to identify a No-Observed-Effect-Level (NOEL) in the study population. Studies of this type need to have sufficient statistical power to detect biologically significant differences between controls and treated subjects.

However, it is critical to distinguish between statistical significance and biological significance of observed events. To label an observed change as "statistically significant" merely says there is sufficient evidence that it is not an artifact of random variation. Any difference, no matter how small or how biologically unimportant, can be statistically significant if the sample sizes are large enough or the variability small enough. Statistical significance reflects the subject and measurement variability and the amount of experimental resources devoted to the experiment. While statistical significance gives some assurance that an effect is really present, it says nothing about the toxicological or biological importance of that effect.

Biological significance implies that an observed event has important toxicological consequences that are relevant to the particular issue being considered. An increase or decrease in the level of a specific endpoint in the population may not be of biological significance. To be biologically significant, such an average change superimposed upon existing individual and temporal variation should have a toxicologically meaningful impact on a relevant portion of the population.

Statistical and biological significance are unrelated. It is not appropriate to use statistical significance to imply biological significance. A statistical test might be sensitive to small differences that are not biologically significant. On the other hand, a poorly designed study might be unable to detect a truly important biological difference. If simple statistical hypothesis tests are used, they should be designed so that their power to detect biologically relevant effects is adequate. The level of red blood cell (RBC) or plasma cholinesterase inhibition that is biologically significant as a biomarker for a potentially adverse human health effect is generally believed to be in the range of 15 to 25%.

The information on NOELs from a human study is combined with other information in the human risk characterization. That other information is used to characterize inter-individual variability within the study population and to extrapolate from the study population to general populations. Thus, during the human risk characterization, the data collected from human studies to characterize the NOEL in the study population is combined with other information to identify individual differences in sensitivity and potential sensitive subpopulations. As necessary, appropriate safety factors are then added to account for intra-human variability and to extrapolate from adults to children.

The Variation of RBC and Plasma Cholinesterase Levels in Human Cholinesterase Studies

As is shown in the Appendix, the critical first step in the evaluation of power is the determination of the within-subject variation in cholinesterase (ChE) levels. For a specified within-subject variation, the power of common study designs to overcome this variation can be assessed. RBC and plasma cholinesterase (ChE) results from nine different human studies were supplied to JSC Sielken for evaluation. The characteristics of these nine studies are listed in Table 1. Data for RBC ChE levels were available for all studies. In addition, plasma ChE levels were available for seven of the nine studies. For the most part, the baseline cholinesterase data were in the placebo groups. In some cases, however, pre-dosing ChE level data from treated subjects were also available to augment the analysis of placebo subjects.

Estimates of the within-subject coefficient of variation, CV_E, were obtained from all studies. A separate variability analysis of the data was conducted for each gender and ChE type in each study. If there were pre-dosing treatment groups available for a study, an additional analysis was conducted on these subjects as well.

The ability to estimate variance components from a linear model is a common feature of most general statistical packages such as SAS®, S-Plus®, etc. We employed the general linear model analysis platform in JMP® software. The value of CV_E was computed as the estimated within-subject standard deviation as a percent of the overall mean ChE level. These estimates are all listed in Table 2.

It is clear from the results in Table 2 that the estimated within-subject variation for plasma ChE is smaller than that for RBC. This is not unexpected given the complexity of the ChE analytical method for RBC relative to that for blood plasma. The median CV_E for plasma ChE appears to be approximately 5.5% for both sexes. For RBC, the median CV_E is 6.5% for females and 8.5% for males. A smaller CV_E for females implies greater power for females.

Power Assessment for Human Cholinesterase Studies

Variability analysis of the untreated subjects in the nine ChE studies in Table 1 indicates that within-subject variation CV_E is likely to be in the range of 5.5% for plasma ChE to as high as 8.5% for RBC ChE. The corresponding power to detect ChE inhibition depends on the particular study designs that are commonly used in practice to overcome this variation. Table 3 summarizes the essential components of the study designs used in the nine studies providing the variability information. These components are listed separately for the female and the male portions of each study. The major components that characterize the design are as follows:

- Total number of time points available for analysis (m), which excludes screening values,

- Number of available time points used to establish a baseline mean (b),
- Number of subjects in the placebo group (n_0),
- Number of subjects per treated group (n_i), and
- Number of treated or 'dose' groups (c).

As shown in the Appendix, increasing m, b, c, n_0, and n_i increases the power.

Increasing m, b, c, n_0, and n_i indirectly increases the power by increasing the degrees of freedom (ν). As ν increases, the Student's t distribution approaches the normal distribution. Once the value of ν exceeds 30, there is very little further indirect increase in the power. As can be seen from Table 3, all of the designs yield degrees of freedom that exceed 100.

The number of baseline points (b) and the number of subjects per group (n_0 and n_i) directly increase power. Increasing b increases the power, but the amount of increase in the power decreases as b increases. For moderate to large values of b, say b equal to 6 or more, the power is dominated by the number of subjects per group (n_0 and n_i).

In the study designs in Table 3, the number of human subjects in the placebo group ranges from 3 to 12 subjects. The number of subjects per treated group ranges from 3 to 7. For each combination of n_0 and n_i in these ranges, Table 4 gives the computed inhibition levels that are detectable with 80% power. Each combination of group sizes gives a range of detectable inhibition values. This range corresponds to the range of CV_E values estimated in the previous section. It is clear from Table 4 that all studies are able to detect ChE inhibition of 15%-20% or more with 80% power, even at higher CV_E values. Many combinations can detect ChE inhibitions below 10% with 80% power.

In Table 4, the number of time points (m) and the number of baseline points (b) were assumed to be equal to 16 and 6, respectively. These are the values in the recent human studies described in Table 1. In addition, it was assumed that there was only a single dose group (c=1). (This latter assumption minimizes the degrees of freedom; however, if the degrees of freedom value was less than 100, it was set equal to 100 because all of the values in Table 3 exceeded 100.)

For simplicity in presentation, the results in Table 4 made assumptions about the number of baseline and total time points used. It is of interest to determine what the range of powers would be for the exact study configurations employed in the studies in Table 1. These powers are shown in Tables 5-7 for within-subject coefficients of variation of 8.5% and 5.5%. Table 5 shows the powers to detect a true inhibition level of 15%. Tables 6 and 7 show the powers to detect inhibitions of 20% and 25%, respectively.

Any seemingly inconsistent results between the detectable levels from Table 4 and the powers in Tables 5-7 are primarily due to the effect of the number of baseline time points. The results in Table 4 assumed the most common value for b; that is, b=6.

The powers to detect a 15% inhibition level in Table 5 exceed 80% in all but a few studies under the worst-case assumption for the within-subject variability (CV_E =8.5%). With 5.5% variability, every study has at least 95% power to detect 15% inhibition. Even at the highest level of variation (8.5%), every study has better than 90% power to detect 20% ChE inhibition (Table 6). When the within-subject CV_E is 5.5%, all powers exceed 99% in Table 6. Lastly, for inhibition levels at or exceeding 25%, the detection certainty is near 100% for every study.

Conclusions

The analyses in this report demonstrate that the experimental designs currently employed for human RBC and plasma ChE evaluation are all quite powerful for inhibition levels in the range of 15-25% below placebo levels.

It should be noted that additional procedures could be used to increase the power above the 'baseline' power computed here. For example, it is frequently possible to combine sexes or dose groups when appropriate to increase the effective number of subjects and, hence, the power. In addition, many testing laboratories have additional placebo subjects available from companion or similar studies that could be used to augment the power. For example, several of the studies we examined used 3 female placebo subjects each. Adding 3 female placebo subjects from each of 3 other studies conducted contemporaneously at the same laboratory would increase the effective number of placebo subjects used from 3 to 12. As is shown in the last column of Table 4, even under the larger variability assumption (8.5%), this increase in the number of subjects in the placebo group pushes the level of detectable inhibition (at 80% power) from 16% down to 11%.

Based on the power results from this analysis, statistical comparison of cholinesterase levels between any dose group and controls in recent human studies should have little difficulty in detecting a 15-25% inhibition rate if it were present. The magnitude of blood ChE inhibition that is biologically significant as a biomarker for a potentially adverse human health effect is generally believed to be of the order of 15-25%. The sizes of the human studies investigated for this analysis generally have power in excess of 80% to detect these levels of cholinesterase inhibition.

Table 1.
List of Studies Used in This Analysis

Study Code	ChE Data Provided	Subject Group	Pre Dose Times	Post Dose Times	Sex	Number of Subjects
A	RBC, Plasma	Placebo	2	13	Males	4
B	RBC, Plasma	Placebo	3	14	Males	10
C	RBC, Plasma	Placebo	6	10	Males	11
					Females	3
		Treated	6	0	Males	27
					Females	7
D	RBC	Placebo	2	12	Males	6
					Females	6
E	RBC	Placebo	2	12	Males	6
					Females	6
F	RBC, Plasma	Placebo	6	10	Males	12
					Females	3
		Treated	6	0	Males	28
					Females	7
G	RBC, Plasma	Placebo	6	10	Males	12
					Females	3
		Treated	6	0	Males	28
					Females	7
H	RBC, Plasma	Placebo	6	10	Males	9
					Females	6
		Treated	6	0	Males	18
					Females	12
I	RBC, Plasma	Placebo	6	10	Males	10
					Females	3
		Treated	6	0	Males	18
					Females	12

Table 2.
Estimated Within-Subject Percent Coefficient of Variation (CV_E) for RBC and Plasma Cholinesterase

	Estimated Coefficient of Variation, % (No. of Subjects)			
	RBC		Plasma	
Study Code	**Males**	**Females**	**Males**	**Females**
A	9.6 (4)	–	5.2 (4)	–
B	8.8 (10)	–	5.5 (10)	–
C	9.3 (11) 8.1 (27) [1]	6.6 (3) 6.4 (7)	7.7 (11) [2] 7.4 (27)	9.9 (3) 6.4 (7)
D	4.1 (6)	3.6 (6)	–	–
E	3.6 (6)	2.9 (6)	–	–
F	8.7 (12) 10.2 (28)	6.5 (3) 8.7 (7)	5.8 (12) 5.2 (28)	3.8 (3) 6.7 (7)
G	7.5 (12) 8.2 (28)	6.1 (3) 7.2 (7)	5.0 (12) 5.8 (28)	5.4 (3) 4.2 (7)
H	9.0 (9) 8.0 (18)	13.0 (6) 8.3 (12)	5.8 (9) 5.6 (18)	5.8 (6) 6.4 (12)
I	7.8 (10) 6.6 (28)	5.8 (3) 4.0 (7)	4.8 (10) 5.6 (28)	4.5 (3) 6.0 (7)
Median [3]	8.5	6.5	5.5	5.3

[1] If present, the 2nd value is that obtained from pre-dose time points of treated subjects.

[2] This value is inflated by the presence of 3 anomalous subject-time point combinations. Removal of these 3 values would reduce the estimated CV_E to 5.4%. This would have no effect, however, on the median value over male plasma studies.

[3] Median of study averages

Table 3.
The Array of Study Design Configurations Used in the Nine ChE Human Studies Provided for Variability Analysis

Study Code	Total No. Time Points	No. Points Used as Baseline	Subjects in Placebo Group	Subjects per Dose Group	No. of Dose Groups	Degrees of Freedom[1]
Designs Used for Males						
A	15	2	4	5	3	210
B	18	3	10	5	6	561
C[2]	16	6	11	3	2	480
	16	6	11	7	3	480
D	14	2	6	6	2	195
E	14	2	6	6	1	130
F	16	6	12	7	4	525
G	16	6	12	7	4	525
H	16	6	9	6	3	345
I	16	6	10	6	3	360
Designs Used for Females						
C	16	6	3	7	1	120
D	14	2	6	6	2	195
E	14	2	6	6	1	130
F	16	6	3	7	1	120
G	16	6	3	7	1	120
H	16	6	6	6	2	225
I	16	6	3	6	1	105

[1] Degrees of freedom associated with a repeated measures estimate of the standard error for mean inhibition. See formula in Appendix.

[2] In study C, for males, there were 3 subjects/dose for the lowest 2 doses and 7 subjects/dose for the highest 3 doses.

Table 4.
True Percent Inhibition Detectable with 80% Power for Range of Within-Subject Coefficient of Variation Between 5.5% and 8.5%[*]

Number of Subjects in Placebo Group	Number of Subjects per Dose Group				
	3	**4**	**5**	**6**	**7**
3	12-19	11-18	11-17	11-16	10-16
4	11-18	11-16	10-15	10-15	9-14
5	11-17	10-15	9-15	9-14	9-13
6	11-16	10-15	9-14	9-13	8-13
7	10-16	9-14	9-13	8-13	8-12
8	10-16	9-14	8-13	8-12	8-12
9	10-15	9-14	8-13	8-12	7-12
10	10-15	9-14	8-13	8-12	7-11
11	10-15	9-13	8-12	8-12	7-11
12	10-15	9-13	8-12	7-11	7-11

[*] The first percent difference in the range corresponds to a within-subject coefficient of variation (CV_E) of 5.5% and the second to a CV_E of 8.5%. These calculations assume that 16 total time points are available and that 6 are used to establish a baseline mean. Minimum degrees of freedom were set to 100.

Table 5.
The Power to Detect a True Percent Inhibition Level of 15% Computed for the Study Design Configurations Used in the Nine ChE Human Studies Provided for Power Analysis

Study Code	Total No. Time Points	No. Points Used as Baseline	Subjects in Placebo Group	Subjects per Dose Group	No. of Dose Groups	% Power Assuming a CV* of 8.5%	5.5%
Designs Used for Males							
A	15	2	4	5	3	69	95
B	18	3	10	5	6	87	>99
C [2]	16	6	11	3	2	80	99
	16	6	11	7	3	96	>99
D	14	2	6	6	2	80	99
E	14	2	6	6	1	80	99
F	16	6	12	7	4	96	>99
G	16	6	12	7	4	96	>99
H	16	6	9	6	3	93	>99
I	16	6	10	6	3	93	>99
Designs Used for Females							
C	16	6	3	7	1	76	98
D	14	2	6	6	2	80	99
E	14	2	6	6	1	80	99
F	16	6	3	7	1	76	98
G	16	6	3	7	1	76	98
H	16	6	6	6	2	88	>99
I	16	6	3	6	1	74	97

* The within-subject coefficient of variation, CV_E

Table 6.
The Power to Detect a True Percent Inhibition Level of 20% Computed for the Study Design Configurations Used in the Nine ChE Human Studies Provided for Power Analysis

Study Code	Total No. Time Points	No. Points Used as Baseline	Subjects in Placebo Group	Subjects per Dose Group	No. of Dose Groups	% Power Assuming a CV* of 8.5%	5.5%
Designs Used for Males							
A	15	2	4	5	3	89	>99
B	18	3	10	5	6	98	>99
C [2]	16	6	11	3	2	95	>99
	16	6	11	7	3	>99	>99
D	14	2	6	6	2	95	>99
E	14	2	6	6	1	95	>99
F	16	6	12	7	4	>99	>99
G	16	6	12	7	4	>99	>99
H	16	6	9	6	3	99	>99
I	16	6	10	6	3	99	>99
Designs Used for Females							
C	16	6	3	7	1	93	>99
D	14	2	6	6	2	95	>99
E	14	2	6	6	1	95	>99
F	16	6	3	7	1	93	>99
G	16	6	3	7	1	93	>99
H	16	6	6	6	2	98	>99
I	16	6	3	6	1	92	>99

* The within-subject coefficient of variation, CV_E.

Table 7.
The Power to Detect a True Percent Inhibition Level of 25% Computed for the Study Design Configurations Used in the Nine ChE Human Studies Provided for Power Analysis

Study Code	Total No. Time Points	No. Points Used as Baseline	Subjects in Placebo Group	Subjects per Dose Group	No. of Dose Groups	% Power Assuming a CV* of 8.5%	5.5%
Designs Used for Males							
A	15	2	4	5	3	97	>99
B	18	3	10	5	6	>99	>99
C [2]	16	6	11	3	2	99	>99
	16	6	11	7	3	>99	>99
D	14	2	6	6	2	99	>99
E	14	2	6	6	1	99	>99
F	16	6	12	7	4	>99	>99
G	16	6	12	7	4	>99	>99
H	16	6	9	6	3	>99	>99
I	16	6	10	6	3	>99	>99
Designs Used for Females							
C	16	6	3	7	1	99	>99
D	14	2	6	6	2	99	>99
E	14	2	6	6	1	99	>99
F	16	6	3	7	1	99	>99
G	16	6	3	7	1	99	>99
H	16	6	6	6	2	>99	>99
I	16	6	3	6	1	98	>99

* The within-subject coefficient of variation, CV_E

Appendix

The Development of Power Formula for ChE Studies

Most cholinesterase (ChE) inhibition studies using human subjects have the same basic form. Subjects are randomly assigned to a placebo group and c treated groups. The number of subjects in each of the c+1 dose groups may or may not be the same. Let n_i denote the number of subjects in the i-th dose group. Excluding screening measurements, each subject has ChE measurements taken at m fixed time points during the course of the study. Of these m separate measurements, b of them will be taken before dosing begins and averaged to establish a baseline ChE value for the subject. In general, the primary interest is in detecting substantial levels of inhibition (i.e., decrease) from baseline in the ChE values at one or more post-dosing time points. Although this general design is followed by most studies, the specific values of m, b, n_i, and c vary somewhat from study to study.

The determination of power requires knowledge of the distribution of the test statistic that will be used to determine statistical significance. In this paper we will develop formulas for the power to detect specific ChE inhibition levels assuming a simple repeated measures approach to the analysis of the general design above. Although other statistical methodologies are certainly possible, we feel that their impact on power will be minimal.

The general statistical model assumed here is

$$Y_{ijk} = \mu_{ik} + A_{ij} + E_{ijk} . \qquad (1)$$

In this mixed linear model, Y_{ijk} is the ChE result (international units per liter, iu/L) obtained in the study for subject j in dose group i at time point k. The quantity μ_{ik} represents any potential patterns or trends in the ChE data over time shared by all subjects in the dose group including time-dependent chemical effects and any average differences between dose groups. The μ_{ik}'s are constants herein. The last two quantities in (1) are random effects. The quantity A_{ij} is the random effect of subject (ij) on ChE levels at every time point. It represents the subject-to-subject variation in average ChE levels. The effect E_{ijk} represents the random differences in ChE levels within each subject that are not explained by the combination of general trend (μ_{ik}) and subject average (A_{ij}). We assume that all of the A_{ij} are independent and follow a normal distribution with mean zero and variance V_A. Similarly, we assume that the E_{ijk} are all independent and normally distributed with mean zero and variance V_E. This defines the common compound-symmetric repeated measures model. More complex features, such as a correlation structure among the E_{ijk}, could also be included in model (1). We feel, however, that such models are unnecessarily complicated and provide little additional information for the purpose of this investigation of power.

Under model (1), the two variance components, V_A and V_E, completely summarize the entire variation structure in the data. The total variation expected in ChE values would be the sum of the two variance components, V_A and V_E ,

$$V_T = V_A + V_E \, . \qquad (2)$$

As will be shown below, other functions of these components are more directly useful and intuitive. These are the within-subject and total standard deviations, σ_E and σ_T, respectively. These are the square roots of the corresponding variances, V_E and V_T. In the context of the power analysis, it is more intuitive to express each of these two standard deviations as a percent of some mean ChE level. That is,

$$\begin{aligned} CV_E &= (\sigma_E / \text{mean}) \quad 100\% \\ &\text{and} \\ CV_T &= (\sigma_T / \text{mean}) \quad 100\% \, . \end{aligned} \qquad (3)$$

CV_E and CV_T are called relative standard deviations or coefficients of variation (or simply CV's). As will be shown below, the quantity CV_E will be the most critical variation parameter used in power calculations.

In ChE dosing studies some of the time points will be used to establish a baseline ChE level, B_{ij}, for each subject. Letting b be the number of time points used for the background, B_{ij}, is computed as

$$B_{ij} = (1/b) \ \Sigma_k(Y_{ijk}) \qquad (4)$$

where the sum is over k = 1, 2, ..., b. To find B_{ij} in terms of the random effects, (4) can be combined with (1) to get

$$B_{ij} = \mu_{iB} + A_{ij} + E_{ijB} = \mu_B + A_{ij} + E_{ijB} \, . \qquad (5)$$

The quantity μ_{iB} is the mean of all b of the μ_{ik} with k **,** b and E_{ijB} is the mean of the b values of E_{ijk} with k **,** b. We assume that there are no chemical effects prior to dosing and that the selection of subjects is unbiased. In this case, all μ_{iB} should be the same value for all dose groups (i.e., $\mu_{iB} = \mu_B$ for all i). The variance of the B_{ij} is just the sum of the variance of A_{ij} and the variance of E_{ijB}. The variance of A_{ij} is just V_A. Because E_{ijB} is a mean of b values, E_{ijB} has a variance reduced to V_E/b. Thus,

$$V(B_{ij}) = V_A + V_E/b \, . \qquad (6)$$

For each post-treatment time point, k > b, the ChE values are next expressed as a difference from the baseline

$$L_{ijk} = Y_{ijk} - B_{ij} \, . \qquad (7)$$

In terms of (1) and (5), this difference is

$$L_{ijk} = \mu_{ik} - \mu_B + E_{ijk} - E_{ijB} \,. \tag{8}$$

Note that the subject-to-subject random effect A_{ij} is completely absent from the difference from the baseline. Because the μ_{ik} and μ_B are constants, the variance of L_{ijk} is just the sum of the variance of the random effects in (8); i.e.,

$$V(L_{ijk}) = V(E_{ijk}) + V(E_{ijB}) \;=\; V_E \;\; (1+1/b) \,. \tag{9}$$

In general, because V_A is often quite substantial, there is a great reduction in the variance of the difference, L_{ijk}, compared with the variation in the raw ChE values, Y_{ijk}. As seen from (9), the variance can be reduced, to a point, by using a larger number of time points to establish a baseline mean.

It should be mentioned that many ChE studies use a percent difference from the baseline, (L_{ijk}/B_{ij}) 100%, rather than the simple difference defined in (7). When the baseline value is in the denominator of this statistic, the between-subject variance does not cancel out completely as it does with L_{ijk}. However, for small true percent deviations (i.e., small μ_{ik} - μ_B compared to μ_B) and large b, the impact of the between-subject variation on this statistic is minimal.

For each post-treatment time point, $k > b$, the mean of the difference L_{ijk} over the n_i subjects in the dose group is

$$L_{iHk} \;=\; (1/n_i) \;\; \Sigma_j(L_{ijk}) = \mu_{ik}\text{-}\mu_B + E_{iHk} - E_{iHB} \tag{10}$$

and its variance is

$$V(L_{iHk}) \;=\; V(E_{iHk}) + V(E_{iHB}) \;=\; V_E \;\; (1+1/b) \;\; (1/n_i) \,. \tag{11}$$

In most cases, it is not adequate to merely test if the expected value of L_{iHk} is equal to zero for a dose group. Even for the placebo group the expected deviation μ_{0k}-μ_B will not usually be zero, and, hence, a ‘placebo effect’ on observations at post-treatment times compared to pre-treatment times would confound any test for dose effects. The preferred, and more typical, approach is to compute ChE ‘inhibition’ as the difference between the placebo and dose group means. That is,

$$D_{ik} \;=\; L_{0Hk} - L_{iHk} \;=\; \mu_{0k} - \mu_{ik} \; + E_{0Hk} - E_{0HB} + E_{iHk} - E_{iHB} \,. \tag{12}$$

The variance of D_{ik} is then

$$V(D_{ik}) = V(E_{0Hk}) + V(E_{0HB}) + V(E_{iHk}) + V(E_{iHB})$$

$$= V_E \ (1/n_i + 1/n_0) \ (1+1/b) \,. \qquad (13)$$

The standard error of this difference, σ_D, is simply the square root of the variance. If an estimate of this standard error, say S_D, is available from the study data, then a t-statistic is used to test the hypothesis that the expected inhibition in ChE level at post-treatment time k, $\Delta_{ik} = \mu_{0k} - \mu_{ik}$, is zero. This test statistic is the ratio

$$t = D_{ik} / S_D \,. \qquad (14)$$

When the average inhibition due to the chemical, Δ_{ik}, is actually zero, the test statistic t will follow a Student's t distribution with ν degrees of freedom. The value of ν will depend upon how many independent 'pieces' of information are used in getting the estimate S_D from the study data on V_E and equation (13). For the simple repeated measures designs typically employed, the number of degrees of freedom will be

$$\nu = (m-1) \ \ \Sigma_i (n_i - 1) \qquad (15)$$

where the sum is over all c+1 dose groups.

Standard statistical methodology would compare the computed value of the test statistic t in (14) to a critical value, the 100(1-α)-th percentile ($T_{1-\alpha}$) of the Student's t distribution with ν degrees of freedom. The quantity α is the specified probability of falsely finding a (statistically) significant difference. If the test statistic t is greater than $T_{1-\alpha}$, then this is taken as evidence that there is an actual inhibition of ChE levels in the dose group compared with those in the placebo group. (Statistical significance would not, however, mean that the true inhibition is exactly the observed difference D_{ik}.)

The power is the probability that the test will give a 'significant' result (i.e., $t > T_{1-\alpha}$) when the actual difference in ChE is some specified value Δ. That is,

$$\text{Power} = P\Delta = \text{Prob}\ (\ t > T_{1-\alpha} \mid \text{actual inhibition is } \Delta\ ) \,. \qquad (16)$$

To find the power for any true inhibition level, it is necessary to know the distribution of the test statistic t for different values of Δ. Only when $\Delta=0$ does the test statistic t follow the well-known Student's t distribution. When $\Delta>0$, the test statistic t follows a non-central t distribution. This is more complicated than the simple t distribution and depends not only on degrees of freedom, ν, but also on a non-centrality parameter, δ. For the type of comparison in (14), the non-centrality parameter is

$$\delta = (\Delta/\sigma_E) \ \ [(1/n_i + 1/n_0) \ \ (1+1/b)]^{-1/2} \,. \qquad (17)$$

Here, σ_E is the actual, not the estimated, within-subject standard deviation, defined as the square root of V_E. Thus, the ratio of the true inhibition to the true within-subject standard deviation is a critical part of the non-centrality parameter. Also, note that, for any specific true mean ChE level,

$$\Delta/\sigma_E = [\ (\Delta/\text{mean})\ 100\%\ ]\ /\ [\ (\sigma_E/\text{mean})\ 100\%\ ]$$

$$= \ \%\Delta\ /\ CV_E\ . \qquad (18)$$

It is generally more intuitively appealing to discuss power in terms of the equivalent percent inhibition (%Δ) and percent within-subject standard deviation (CV_E).

The mathematical form of the non-central t distribution is rather complex but is incorporated in commercial power analysis software such as nQuery Advisor® (from Statistical Solutions). For spreadsheet-oriented power calculations, the non-central t distribution can also be approximated quite closely by the distribution of a Student's t variate plus the value δ. That is,

$$P\Delta = \text{Prob}(\ t > T_{1-\alpha}\ |\ \Delta > 0\ ) \approx \text{Prob}(\ t + \delta > T_{1-\alpha}\ |\ \Delta = 0\ )\ . \qquad (19)$$

This means PΔ can be found just by computing δ and finding

$$P\Delta = \text{Prob}(\ t\ > T_{1-\alpha}\ - \delta\ |\ \Delta = 0\ ) \qquad (20)$$

from the Student's t distribution using readily available tables or utility functions available in spreadsheet software. Note from (20) that the power will increase as δ increases. This means that power increases as the true ChE inhibition (or percent inhibition) or the number of subjects increase. Also, the power increases as the within-subject variation, σ_E, gets smaller.

JMPR does not view plasma cholinesterase inhibition as a relevant toxi